UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

IN RE:
 	
OPENAI, INC.,
COPYRIGHT INFRINGEMENT LITIGATION
 	
This Document Relates To:
 	
     Case No. 1:23-cv-08292
     Case No. 1:23-cv-10211
     Case No. 1:24-cv-00084
     Case No. 1:25-cv-03291
     Case No. 1:25-cv-03297
     Case No. 1:25-cv-03482
     Case No. 1:25-cv-03483
	Case No. 25-md-3143 (SHS) (OTW)

OPENAI DEFENDANTS' REPLY IN SUPPORT OF MOTION TO STRIKE

TABLE OF CONTENTS

TABLE OF AUTHORITIES

ii

I.   INTRODUCTION

Before the Consolidated Class Action Complaint ("Complaint") was filed, in Plaintiffs' own words, "[t]his case [was] about whether OpenAI's reproduction of pirated books to train its large language models [constitutes] copyright infringement."[1] The Court's order on the scope of the Complaint was clear: it should "include simply the products and causes of action that have already been asserted."[2]

Plaintiffs' Complaint violates this directive in two ways. The Complaint materially expands the case by adding new large language models, including models that did not exist at the time any prior complaint was filed and one that was released ***after*** OpenAI filed the instant Motion. The Complaint also advances a brand new infringement claim as to any work downloaded but never used for model training. These models and the "download only" claim would require a "never-ending process" for discovery that would substantially delay the Court's fair use decision.[3]

In support of their effort to add models to the case, Plaintiffs mischaracterize the pleadings and the scope of discovery in the centralized *Tremblay* action[4], neither of which supports their argument. Likewise, Plaintiffs do not and cannot dispute that adding these models would dramatically expand discovery and delay proceedings.

For their new "download only" claim, Plaintiffs argue the download of text-based works has been at issue because Plaintiffs previously alleged OpenAI downloaded their works and ***then***

---

[1] *Authors Guild*, Dkt. 194; *see also id.*, Dkt. 44, 11/29/2023 Hearing Tr. at 11:21–12:1 ("We're focusing on the mass copying of protected works for purposes of training the large language model.").

[2] 5/22/2025 Hearing Tr. at 37:17-19.

[3] *Id.* at 25:24-25 ("If I'm going to allow amended claims, new claims, new products, it would be a never-ending process.").

[4] *In re OpenAI ChatGPT Litigation,* No. 3:23-cv-03223 (N.D. Cal.) ("*Tremblay*").

used them for model training. Plaintiffs miss the point: this Complaint introduces a new copyright claim based upon downloading ***untethered to training***. ***Every*** complaint and ***every*** class definition prior to the amended Complaint alleged that OpenAI was liable for using Plaintiffs' works ***in training***. ***No*** complaint or class definition previously alleged a "download only" claim. OpenAI's Motion to Strike only asks the Court to block Plaintiffs' late-stage addition of that claim.

The Court admonished the parties that it "ha[d] no current intention of changing" the August 14, 2026 summary judgment motion deadline, which is "where this [case] is going."[5] That deadline cannot be met if Plaintiffs expand the scope of at-issue models or their claims. To move the case forward efficiently, the Court should strike Plaintiffs' allegations concerning out-of-scope models and the "download only" infringement claim.

II.   ARGUMENT

A.   The Court Should Strike Allegations Regarding GPT-4V(ision), GPT-4.5, GPT-5, and Model "Derivatives" and "Successors."

The Court directed Plaintiffs to limit the Complaint to "the products . . . that have already been asserted," meaning the "explicit models laid out" in OpenAI's supplemental response to Interrogatory 11. 5/22/2025 Hearing Tr. at 37:15-19; 45:13-16. Attempting to shoehorn models into the Complaint, Plaintiffs claim that other models not laid out in that interrogatory response: (a) were previously alleged in *Tremblay*, (b) were "involved" in discovery there, and (c) should be addressed now for efficiency's sake. Opp. at 23-24. Plaintiffs misrepresent the *Tremblay* complaint, overstate the scope of discovery in *Tremblay*, and make no attempt to address the discovery burden—and its impact on scheduling—their amendments would impose.

---

[5] 6/26/2025 Hearing Tr. at 291:19-291:4.

First, GPT-4V(ision), GPT-4.5, GPT-5, and all models' "derivatives" and "successors" were **not** within the scope of *Tremblay*. To suggest otherwise, Plaintiffs point to one paragraph in the *Tremblay* complaint discussing OpenAI's model development history. *See* Opp. at 23 (citing *Tremblay,* Dkt. 120 ¶ 36). But Plaintiffs never alleged their works were used to train GPT-4V(ision), GPT-4.5, GPT-5, or any model's "derivatives" or "successors." Rather, Plaintiffs' copyright claim focused on specific models made available to customers: GPT-3, GPT-3.5, GPT-3.5 Turbo, GPT-4 and GPT-4 Turbo. *See, e.g.*, *Tremblay* Dkt. 120 ¶¶ 40-51; *Tremblay* Dkt. 217 at 3 n.2. Plaintiffs' passing reference to further models does not justify discovery on **every** model indefinitely.

Second, Plaintiffs argue the scope of discovery in *Tremblay* was broader than in the New York actions because it "involved" limited discovery into GPT-4.5 and other then-in-development models. According to Plaintiffs, this limited discovery makes every model fair game for full discovery. But Plaintiffs mischaracterize the *Tremblay* record. Plaintiffs cite Judge Illman's order for the proposition that "discovery in *Tremblay* involved [the further-asserted models]." Opp. at 23. But the order says otherwise. The *Tremblay* Plaintiffs sought full discovery "on **all** GPT-class models, including in-development models," as they seek now. *Tremblay* Dkt. 217 at 2 (emphasis added). *Judge Illman rejected Plaintiffs' attempt* and ordered a narrower request. 12/17/2024 Hearing Tr. at 51:22-23. OpenAI then made a compromise offer (i) to produce **only** the **pre-training** data for GPT-4.5 and (ii) to "not exclude" documents otherwise responsive to a limited and specific set of Plaintiffs' requests solely because they concern "model[s] in development." *Tremblay* Dkt. 246 at 5. Judge Illman adopted OpenAI's compromise and did **not** order full discovery into GPT-4.5 or other in-

development or future models or say anything about model "derivatives." *Tremblay* Dkt. 247 at

3-4. Indeed, Judge Illman **barred** full discovery into new models. *Id.*

       Third, as the Court warned when it limited Plaintiffs to the models already at issue, the

addition of an unknown number of "derivative" and "successor" models developed or released

would delay a fair use decision. Specifically, these models "would add a minimum of four to six

months of discovery to undertake additional investigations, add new custodians, and collect and

make available an immense volume of custodial and non-custodial documents and training data."

Mot. at 11. New and distinct capabilities of additional models will also require unique expert

discovery. Plaintiffs fail to address this delay or provide authority for their proposed case

expansion. As the Court observed, the consolidated class action is "leading to a fair use opinion"

and amendments adding "products and causes of action [beyond those] that have already been

asserted" "are not getting [the parties] there." 5/22/2025 Hearing Tr. at 37:12-19. The Court

"[may] handle any tagalong actions," which will proceed efficiently with the parties informed by

the Court's forthcoming fair use decision. *Id.* at 37:21-23. That approach is "a more efficient

way to resolve it" than to allow for "amendments and amendments and amendments" with every

released model as the case proceeds. *Id.* at 37:20-21, 36:23-24.

       B.   **The Court Should Strike Plaintiffs' New Infringement Claim.**

       The Court also should reject Plaintiffs' fallback request to expand the Complaint to add a

"download only" claim.

       1. Plaintiffs' "Download Only" Infringement Claim is New.

       Plaintiffs' new "download only" claim defies the Court's directive to limit the Complaint

to "claims that already exist in the action." *Id.* at 24:19-20. As Plaintiffs maintained before

filing the Complaint, "[t]his case is about whether OpenAI's reproduction of pirated books to

train its large language models is copyright infringement." *Authors Guild v. OpenAI, Inc.,* No.

1:23-cv-08292, (S.D.N.Y.) ("*Authors Guild*"), Dkt. 194 (emphasis added); *see also id.*, Dkt. 143

("the core factual question of the case [is] [w]hat data did OpenAI use to 'train' its [LLMs]");

id., Dkt. 169 (same). After recognizing that their longstanding infringement claim was at risk in

light of the summary judgment hearings in *Bartz v. Anthropic* and *Kadrey v. Meta*, Plaintiffs

pivoted to the new "download only" claim.[6] Plaintiffs' four arguments attempting to add the

new claim are all wrong on the law or misstate the record.

  First, Plaintiffs conflate longstanding factual allegations about downloads *for training*

with the new "download only" claim. *See* Opp. at 9-10, 17-18. While the prior complaints all

limited their claims to model training, they—of course—alleged facts about how OpenAI

acquired the relevant data (*e.g.*, by downloading) for that training. Those background allegations

will remain if OpenAI prevails on this Motion. *See* Opp. at 17-18. But works downloaded and

not used for training have never been the subject of a claim. ***Every*** prior complaint was focused

on training, with allegations about how works were obtained for that purpose. *See* Mot. at 4-5

(citing earlier allegations concerning downloads ***for training***); Opp. at 8 (similar). But there was

no claim for works ***not used for training*** in any of the ten prior complaints. The Motion to

---

[6] In hearings earlier this year, two courts expressed a clear view that the fair use defense would apply to the use of text-based works for large language model training. *See Bartz v. Anthropic PBC*, No. 3:24-cv-05417 (N.D. Cal.), Dkt. 208, 5/22/2025 Hearing Tr. 40:20-23 ("[T]hat's your clearest point is that they should have paid for the first use instead of stealing it off the Internet. That is your best point. But after that it's not so clear."); *Kadrey v. Meta Platforms, Inc.*, No. 3:23-cv-03417 (N.D. Cal.), Dkt. 590, 5/1/2025 Hearing Tr. 97:1-4 (plaintiff's copyright claim is "taken away by fair use unless the plaintiff can show that the market for their actual copyrighted work is going to be dramatically affected . . . "). Plaintiffs correctly read the tea leaves: both courts held the use of text-based works for training was a fair use. *See Bartz*, 2025 WL 1741691, at *18 (N.D. Cal. Jun. 23, 2025) ("*The copies used to train specific LLMs* were justified as a fair use."); *Kadrey,* 2025 WL 1752484, at *23 (N.D. Cal. Jun. 25, 2025) ("copying these plaintiffs' books for use as LLM training data" was fair use). While the *Bartz* decision—which allowed a claim for works downloaded but not trained on to proceed—explains Plaintiffs' amendment, it cannot change the claims asserted or the scope of discovery in *this* case.

Strike is directed at this expansion, and at the **new** allegations regarding the **new** claim that

OpenAI is liable for the alleged download of works **regardless** of their use for training. *See* Mot.

at 7. Should OpenAI prevail on the instant Motion, Plaintiffs' **new** claim on behalf of "book

copyright holders **whose works were used by Defendants**" would be stricken (Compl. ¶ 14

(emphasis added)) and only Plaintiffs' claim on behalf of writers "whose works were used to

train" would remain. *Authors Guild*, Dkt. 69 ¶ 11.

   Second, Plaintiffs say the new "download-only" allegations were "grounded" in prior

complaints. Opp. at 10 & Table 2. Not so. Prior complaints mentioned downloading works, but

always linked to their use for **training**. *See* Mot. at 5 (**all ten** prior complaints across class cases

brought on behalf of authors whose works were used for training). Plaintiffs added allegations

because courts in the Northern District of California signaled that training claims were deficient

and Plaintiffs' prior complaints did not support the new claim for all works ever "reproduced" or

"maintained" but not ultimately included in a training dataset. Compl. ¶ 302.

   Third, Plaintiffs' argument that their expansion of the class definition is "routine" lacks

record support. *Contra* Opp. at 10-11. Revisions to a class definition must be based on

developments **in the record**. *See Gen. Tel. Co. of Sw. v. Falcon*, 457 U.S. 147, 160 (1982)

("judge[s] remain[] free to modify [class certification orders] in the light of subsequent

developments in the litigation"); *Nayani v. LifeStance Health Grp., Inc.*, No. 22-CV-6833 (JSR),

2023 WL 6311475, at *5 n.5 (S.D.N.Y. Sept. 28, 2023) ("[W]hen the record warrants, it is

appropriate for a district court to modify a proposed class") (quoting *In re SunEdison, Inc. Sec.*

Litig., 329 F.R.D. 124, 135 (S.D.N.Y. 2019)). Here, Plaintiffs seek to broaden the scope of the

class to include countless works OpenAI allegedly acquired from public sources that were not

used for or included in training datasets. Plaintiffs do so without identifying any portion of the record justifying expansion.

Fourth, contrary to Plaintiffs' suggestion, OpenAI never conceded that the earlier *Tremblay* complaint included a "download-only" claim. *Contra* Opp. at 11 ("the only time OpenAI even mentioned downloading in its brief opposing the *Tremblay* motion to amend was to note that downloading was already in the case"). Plaintiffs refer to OpenAI's citation to an allegation in the previously operative complaint. *See Tremblay*, Dkt. 401 at 15 (citing *Tremblay* Dkt. 120 ¶ 58(a)). But the cited allegation—truncated in the Opposition—illustrates exactly the scope of Plaintiffs' longstanding "training" claim. As alleged, OpenAI "violated the copyrights of Plaintiffs . . . when [it] downloaded copies of Plaintiffs' copyrighted books *and used them to train ChatGPT*." (*Tremblay* Dkt. 120 ¶ 58(a) (emphasis added). That claim necessarily excludes works downloaded but never used for training.

2. Plaintiffs' Amendment Would Restart Discovery.

Plaintiffs do not address the additional months it would take to complete discovery on their new "download only" claim or the prejudice to OpenAI. *See* Mot. at 15 (discussing time and resources needed for discovery on Common Crawl acquisition). Instead, Plaintiffs insist that the new claim would not prejudice OpenAI because *all* downloads have always been within the scope of discovery and have been addressed at case hearings. But Plaintiffs (i) misrepresent the course of discovery and (ii) conflate their longstanding training claim with the new "download only" claim.

First, discovery has focused on downloads *for training* and does not address the full scope of Plaintiffs' proposed claim. *Contra* Opp. at 12-15. When Plaintiffs have served discovery concerning the acquisition of data (*e.g.*, downloads), OpenAI has objected to requests for documents and data regarding "books not used in model training" based on the scope of the

7

claims and limited its responses accordingly. *See* Gratz Declaration ¶ 2. OpenAI consistently

applied these limitations throughout discovery. Plaintiffs mischaracterize OpenAI's recent

response to *Tremblay* Interrogatories 21-25, the only discovery response Plaintiffs identify that

was not explicitly limited to "training datasets." In the original interrogatories, Plaintiffs sought

information on *all datasets* (not just training datasets) that contained any copyrighted works,

including a description of (i) how data was obtained, (ii) whether and how data was shared

externally, and (iii) OpenAI's internal use of data. Consistent with its longstanding objection to

discovery on data acquisition untethered to training, OpenAI objected to the overbroad

interrogatories. Plaintiffs then narrowed their interrogatories to seek only "a list of the names of

each dataset that OpenAI downloaded from a known shadow library." *See* Gratz Declaration ¶ 3.

The only responsive information related to the Library Genesis datasets that are the focus of

Plaintiffs' *training* claim; there was no added burden, so rather than engage in motion practice,

subject to its ongoing objections consistent with its approach throughout the case, OpenAI

agreed to confirm what OpenAI had already provided Plaintiffs in discovery. *Id.* That narrow

agreement did not expand discovery, cause delays, or waive OpenAI's Motion to Strike the new

"download only" claim. Plaintiffs cite no other response where OpenAI provided broader

discovery, nor any instance where Plaintiffs sought relief from the Court.

   Most telling is that, on the final day to serve written discovery requests, and more than a

month *after* OpenAI filed this Motion, Plaintiffs attempted to expand discovery by serving *new*

requests focusing on downloaded datasets *not related to LLM training*. *See* Gratz Decl. ¶ 4

(discussing over a dozen document requests and requests for admission specific to text-based

works "not used for training"). Plaintiffs join these "download only" requests with *dozens* more

requests asking for document discovery on *every download from virtually any source*. *Id.*

Based on the new requests, Plaintiffs appear to be preparing ***now***, in the final months of discovery, to develop the ***new*** claim they contend has always been in the case.

Given the focus of the case thus far, discovery would be extended by months, or longer, if Plaintiffs' new "download only" claim is permitted. *See* Mot. at 15-16. Offering no contrary argument, Plaintiffs do not and cannot dispute that the new claim would impose this burden, delay, and prejudice to OpenAI. That is the reason the Court limited the Complaint to claims already asserted, ***denying*** the amendment Plaintiffs seek here. *See* 5/22/2025 Hearing Tr. at 20:17-20; *see also Royal Indem. Co. v. King*, 532 F. Supp. 2d 404, 417 (D. Conn. 2008), *aff'd sub nom. Arrowood Indem. Co. v. King*, 699 F.3d 735 (2d Cir. 2012) (rejecting "new claims [that] would require additional discovery and delay the adjudication of the issues . . . and would require [defendants] to defend an entirely new theory of a case"); *In re Aluminum Warehousing Antitrust Litig.*, No. 13-MD-2481 (KBF), 2016 WL 1629350, at *7 (S.D.N.Y. Apr. 25, 2016) (denying amendments that "effectively change[] [plaintiffs'] claims and proposed class definition . . . in ways that could necessitate substantial new fact and expert discovery").

Second, attorney discussion of data acquisition in the context of model training at hearings and deposition does not mean Plaintiffs' new "download only" claim is within scope. Again, since the beginning, downloads have been at issue ***only*** to the extent relevant to the "the core factual question of the case: [w]hat data did OpenAI use to 'train' its [LLMs]." *Authors Guild*, Dkt. 143 (Plaintiffs' discovery brief). As OpenAI showed, were Plaintiffs' new claim permitted, OpenAI would need to investigate, collect, and produce test material and the raw data that is removed or filtered ***unrelated to*** training. *See* Mot. at 14-15. That expansion would require enormous additional resources to complete discovery and delay a fair use decision.

9

3. Efficient Resolution of Plaintiffs' "Training" Claim Is in the Public Interest.

Limiting Plaintiffs' case to the longstanding "training" claim is the most efficient route toward a fair use decision. Repeating their misstatement that the two separate claims rely on the same facts, Plaintiffs disagree. *See* Opp. at 19-21. Although Plaintiffs have suggested that they "have the facts" to pursue their new "download only" claim, Plaintiffs' discovery conduct disproves their argument. *See* 5/22/2025 Hearing Tr. at 51:25–52:9 ("with literally a couple of months more of discovery . . . we can certainly get to a resolution of the piracy use of [Plaintiffs' works]"). As noted above, Plaintiffs served over a dozen new discovery requests concerning datasets that "did not relate to training" in August 2025, *i.e.*, ***after*** filing the Complaint. *See* Gratz Decl. ¶ 4.

As OpenAI has shown, discovery on that data would take at least multiple months, if not longer. *See* Decl. of Bright Kellogg in Supp. of Mot. to Strike, Dkt. 338 ¶¶ 12-14. The parties should not undertake that challenge now, "slow everything down," and harm the "public interest in these things being resolved." *See* 5/22/2025 Hearing Tr. at 25:16-23, 36:20-24. Instead, as the Court directed, the parties should continue work leading to a fair use decision on the ***existing*** claim. The "download only" claim can be resolved in a tag-along action (if necessary). For instance, the new tag-along action, *Denial v. OpenAI, Inc.*, No. 1:25-cv-06286, (S.D.N.Y.), adopts the "download only" claim among antitrust and other claims, and the parties have agreed a stay is appropriate. *See Denial* Dkt. 1 ¶ 1 (asserting claims based on allegedly "obtain[ing] copies of this data via peer-to-peer file-sharing networks…."). As the Court has observed, that approach is "more efficient." 5/22/2025 Hearing Tr. at 37:20-21.

III.   CONCLUSION

OpenAI respectfully asks the Court to strike Plaintiffs' allegations concerning additional

models (Compl. ¶¶ 5, 81-90, 96, 127, and 161) and Plaintiffs' new copyright claim (*id.* ¶¶ 14,

294, and 302).

Dated: August 28, 2025

/s/ Allison S. Blanco
LATHAM & WATKINS LLP
Andrew M. Gass (*pro hac vice*)
  andrew.gass@lw.com
San Francisco, CA 94111
Telephone: 415.391.0600

Sarang V. Damle
  sy.damle@lw.com
Luke A. Budiardjo
  luke.budiardjo@lw.com
1271 Avenue of the Americas
New York, NY 10020
Telephone: 212.906.1200

Elana Nightingale Dawson (*pro hac vice*)
  elana.nightingaledawson@lw.com
555 Eleventh Street, NW, Suite 1000
Washington, D.C. 20004
Telephone: 202.637.2200

Allison S. Blanco (*pro hac vice*)
  allison.blanco@lw.com
650 Town Center Drive, Suite 20
Costa Mesa, CA 92626
Telephone: 714.540.1235

/s/ Joseph C. Gratz
MORRISON & FOERSTER LLP
Joseph C. Gratz (*pro hac vice*)
  jgratz@mofo.com
Tiffany Cheung (*pro hac vice*)
  tcheung@mofo.com
Caitlin Sinclaire Blythe (*pro hac vice*)
  cblythe@mofo.com
425 Market Street
San Francisco, CA 94105
Telephone: 415.268.7000

Rose S. Lee (*pro hac vice*)
  roselee@mofo.com
707 Wilshire Boulevard, Suite 6000
Los Angeles, CA 90017
Telephone: 213.892.5200

11

/s/ R. James Slaughter
KEKER, VAN NEST & PETERS LLP
Robert A. Van Nest (*pro hac vice*)
  rvannest@keker.com
R. James Slaughter (*pro hac vice*)
  rslaughter@keker.com
Paven Malhotra
  pmalhotra@keker.com
Michelle S. Ybarra (*pro hac vice*)
  mybarra@keker.com
633 Battery St.
San Francisco, CA 94111
Telephone: 415.391.5400

Attorneys for OpenAI

12

CERTIFICATE OF COMPLIANCE

In accordance with Local Civil Rule 7.1(c), I hereby certify that the foregoing Memorandum of Law in support of OpenAI's Motion to Strike is 3,493 words, exclusive of the caption page, table of contents, table of authorities, and signature block. The basis of my knowledge is the word count feature of the word-processing system used to prepare this memorandum.

Dated: August 28, 2025   By: _/s/ Joseph C. Gratz_
                                                    Joseph C. Gratz